University of Zurich[UZH]

Department of Business Administration

UZH Business Working Paper Series

Working Paper No. 374

Dealing with randomness in match outcomes: how to rethink performance evaluation and decision-making in European club football

Marc Brechot and Raphael Flepp

September 2018

http://www.business.uzh.ch/forschung/wps.html

UZH Business Working Paper Series

Contact Details

Marc Brechot

marc.brechot@uzh.ch

Raphael Flepp

raphael.flepp@business.uzh.ch

University of Zurich

Department of Business Administration

Affolternstrasse 56, CH-8050 Zurich, Switzerland

Dealing with randomness in match outcomes: how to rethink performance evaluation and decision making in European club football

Marc Brechot[*] Raphael Flepp[**]

September 2018

Abstract

Football is a low-scoring sport in which a few single moments can change the result of an entire match, regardless of what else happened during the 90 minutes on the field. Thus, random forces can have a substantial influence on match outcomes. However, decision makers in European football clubs often rely heavily on recent match outcomes when evaluating team performance, which can lead to systematic misjudgments. In this paper, we propose a complementary approach for performance evaluation aimed at enabling decision makers to substantially mitigate the tendency to overlook the influence of randomness in match outcomes. We build upon the concept of *expected goals* based on quantified scoring chances and develop a chart that visualizes situations in which a team's true performance throughout a sequence of matches likely deviates from the performance indicated by match outcomes. The insights provided by the chart can systematically alert decision makers of professional football clubs about sensitive situations and should prevent clubs from making flawed decisions when match outcomes are overly biased due to the influence of random forces.

JEL Classification: D81, L83
Keywords: football, expected goals, performance evaluation, decision making

[*]Marc Brechot and Raphael Flepp: Department of Business Administration, University of Zurich.
[**]Corresponding author. Email: raphael.flepp@business.uzh.ch

1 Introduction

European football clubs operate in multi-division league systems on the national level with promotion and relegation between divisions at the end of each season and with opportunities to compete in additional cup competitions. In each season, due to the expectations of owners, fans and the media, as well as for financial reasons, clubs are under pressure to achieve certain seasonal targets, such as winning the league title, gaining promotion, avoiding relegation, progressing in cup competitions or qualifying for a European-level competition for the following season. In this environment, the decision makers for the clubs (e.g., owners and directors) often act with a short-term perspective.

Furthermore, the perspectives of clubs are not only characterized by short-termism but also by a strong outcome focus. This characteristic is perfectly illustrated by former English captain Rio Ferdinand: "It's as simple as that (...) the table never lies, the table is a true marker of where you are supposed to be in the football league" (Henry, 2017, para. 3). Week in, week out, in football leagues across Europe, decisions worth many millions of Euros are based on the logic that match outcomes are intrinsically tied to true performance on the pitch.

However, match outcomes in football are disproportionately influenced by randomness because football is a low-scoring game in which winning and losing is often determined by a single goal. Thus, match results occasionally fail to reflect the true level of play of the two teams on the pitch. Therefore, it is questionable whether match outcomes truly represent a consistent performance indicator, particularly when considering a limited match window inside the scope of a single season. Rather, it must be assumed that systematic misjudgment occurs if outcome-based performance evaluation is applied in situations in which random forces are significant drivers of the results of recent sporting events.

Indeed, the fact that important outcomes are shaped by random forces is difficult to accept because people mistakenly perceive patterns in random sequences (e.g., Henderson, Raynor, & Ahmed, 2012; Taleb, 2005; Tversky & Kahneman, 1974). Moreover, psychological research suggests that the acceptance of random forces as a driver of outcomes becomes even more difficult if the desired outcome is known ex ante, if people take actions for themselves, and if a situation is focused on success (Thompson, 1999). In football, these factors are all highly relevant to match outcomes. Therefore, the decision makers of clubs are likely to exhibit an outcome bias, where they underestimate the role of randomness in match outcomes and assign too much weight to the observed outcomes in their performance evaluation (Baron & Hershey, 1988; Gauriot & Page, 2017). As a result, decision makers fail to make needed adjustments after fortuitous wins and act excessively after unlucky losses (Lefgren, Platt, & Price, 2015).[1]

In this paper, we propose a method for the decision makers of professional football clubs to substantially mitigate the tendency to overlook the influence of randomness in match outcomes. To do so, we draw on the idea of *expected goals* based on quantified scoring chances, which is assessed by public blogs (e.g., Caley, 2015) and professional sports data companies (e.g., OptaPro, 2017a; Prozone Sports, 2015) but in only a few academic papers.[2] The core idea of expected goals based on quantified scoring chances is to give more weight to the actual production process on the pitch instead of relying solely on the scoreline. This approach has several appealing elements. First, scoring chances occur much more often than goals and therefore are less prone to the influence of randomness inherent in single moments of the game. Second,the approach accounts for the fact that different types of scoring chances, for example, a shot taken 5 meters away from the goal versus a shot taken 30 meters away from the goal, are associated with very different probabilities of producing a goal. Third, this approach is football-intuitive. To create as many good chances as possible and to minimize such chances for the opposing team is integral in any reasonable game plan in football. Thus, the concept

[1] This logic also applies if the decision makers base their decisions on the difference between expected match outcomes (e.g., derived from the winning probabilities implied by betting odds) and actual match outcomes because the latter are still subject to randomness.

[2] An early introduction of how to quantify scoring chances goes back to the work of Richard Pollard and his colleagues (Pollard & Reep, 1997; Pollard, Ensum, & Taylor, 2004). More recent work includes Lucey, Bialkowski, Monfort, Carr, and Matthews (2015) and Eggels, van Elk, and Pechenizkiy (2016).

of expected goals based on quantified scoring chances applies to any given playing style and is comprehensible by anyone involved in football.

We use 170,688 shots from all 7,304 matches played in the English Premier League, the French Ligue 1, the German Bundesliga, the Italian Serie A, and the Spanish La Liga in the four seasons from 2013/2014 to 2016/2017 to measure scoring chances and estimate the scoring probability of each shot, which represents the value of a scoring chance. Thereby, we account for the location of the shot on the pitch, the rule setting (i.e., open play, free kick and penalty kick), and the part of the body used. The individual scoring probabilities of multiple shots are then summed over one or multiple matches to derive a cumulative chance value, which is generally referred to as expected goals (e.g., OptaPro, 2017b; Caley, 2015). For example, if a team had three shots in a match, one within the six-yard-box with an estimated scoring probability of 0.40, one from around the penalty spot with an estimated scoring probability of 0.10, and one from far away of the goal with an estimated scoring probability of 0.01, the team has generated chances worth 0.51 expected goals.[3]

To obtain an evaluation measure of performance at the team-level based on quantified scoring chances, we calculate the difference between expected goals created and expected goals allowed by the team across different match samples, ranging from single matches to all the matches in a full season. This measure is football-intuitive because it simply reflects the cumulative value of all the chances created and allowed by a team at a given point in the season. Moreover, the measure should exhibit less random variation than do the plain match results because it depends less on the few moments that led to actual goals scored and conceded but instead considers a much larger part of the team's production on the pitch.

Indeed, we show that the recent difference between the expected goals created and allowed by a team predicts sporting results in the future better than do the past match

[3] The most natural interpretation of 0.51 expected goals is that if the same match with exactly the same shots was played over and over again, the team would be expected to score 0.51 goals on average. Importantly, this figure reflects the three shots that actually have been taken by the team in that match. Thus, there are real goal chances behind expected goal values and not just expectations.

outcomes of a team. Specifically, we compare the goodness-of-fit measure R^2 from a univariate linear regression of the future number of points won on the number of points won from past matches to the R^2 resulting from the regression of the future number of points won on the difference between expected goals created and allowed from past matches. For any combination of the number of previous matches and the number of following matches within the horizon of a full season, the R^2 obtained by including the difference between expected goals created and allowed in the regression is higher than the R^2 obtained by including the number of points won in the regression. A comparison of the R^2 values indicates that expected goals generally contain more information of true performance on the pitch than do match outcomes. Thus, the measure appears to successfully filter out some of the random components that potentially blur match outcomes as a performance evaluation measure. Furthermore, we show that at the individual club level, an overperformance or underperformance of expected goals in terms of actual goals (i.e., match outcomes) is often unsustainable and not due to the qualities of a team that are not captured in our model.

To enable the informational advantage of expected goals to be used to improve the decision making of a club in an as simple way as possible, we construct a chart that visualizes situations where randomness is likely to play a large role in match outcomes.[4] We plot teams' rankings in the official league table during a certain matchweek against their rankings based on the difference between the expected goals created and allowed by the team. If a team is far below the identity line, i.e., the rank in the league table is much lower than the rank based on expected goals, it is suspected that the team is under-rewarded in the league table due to a sequence of overly negative results. By contrast, if a team is far above the identity line, it is suspected that the team is over-rewarded in the league table due to a sequence of overly positive results. In both situations, decision

[4] Of course, such a chart represents only one example of a visualization that can be customized based on the individual preferences and requirements of a particular club.

makers are well-advised to be cautious when drawing conclusions based on a team's rank in the official league table.

Situations in which there are large differences between the two table ranks are those that are most likely to yield decisions based on misjudgments. Thus, we investigate several cases in which large discrepancies between the ranking in the official league table and the ranking based on expected goals arguably led to flawed decision making. For example, the Spanish club Real Betis dismissed Pepe Mel after matchweek 15 of the 2013/2014 season when they were placed last in the official league table but ranked a satisfactory 8th place based on expected goals. Despite replacing the coach, Real Betis was relegated to the 2nd division at the end of the season. Manuel Dominquez Platas, who was a director of Real Betis at the time of Pepe Mel's dismissal, admitted that it had probably been a mistake to sack him and that they focused too much on the unsatisfying match results: "In retrospect, we should have thought about it [his dismissal] a little bit more, but we were last placed for some weeks already." (Lepkowski, 2014, para. 6).

Our paper makes three important contributions. First, we introduce the concept of expected goals based on quantified scoring chances both theoretically and empirically into the sports economics literature and show that such a metric reflects the true performance of teams on the pitch more accurately than do match results. Second, we develop a simple rank comparison chart that alerts decision makers of situations in which random events may have played a crucial role in the club's sporting results. The chart can be easily implemented and understood by decision makers to develop a more reliable picture of a team's true performance on the pitch. Third, we support a new mindset for performance evaluation and decision making of football clubs. Namely, decision makers can complement their existing outcome-based evaluation strategies with more process-oriented and data-driven evaluation strategies. This new mindset can be applied to a wide range of decisions that must be made by professional football clubs, for example, those regarding squad management or the recruitment of coaches.

The remainder of this paper is structured as follows: In Section 2, we develop a theoretical framework of expected goals based on quantified scoring chances. In Section 3, we estimate an expected goal model, and in Section 4, we compare the performance evaluation measures. In Section 5, we address the overperformance and underperformance of expected goals. In Section 6, we focus on the identification of discrepancies between expected goals and match outcomes to improve decision making. In Section 7, we conclude.

2 Expected goals framework

In the production process of football, goals, which ultimately determine who will win and who will lose, are preceded by scoring chances. Thus, the last step in the production process before scoring a goal is to create scoring chances. The concept of expected goals draws on these moments of a football match.

To derive expected goals as a sum of quantified scoring chances, one first must define how scoring chances are identified. The most common approach is to use shots as proxies for scoring chances because shots are direct attempts to score goals and are relatively easy to identify (e.g., Pollard & Reep, 1997; Caley, 2015).[5] Independent of whether a shot translates into a goal, each shot exhibits a certain scoring probability based on its given circumstances. In this section, we develop and discuss eight general factors that are expected to influence the probability of a shot translating into a goal. These factors are listed in Table 1.

The first and probably most obvious factor is the location of the shot on the pitch. Intuitively, a long-range shot from 30 meters must have a low scoring probability because the goalkeeper has sufficient time to react. By contrast, a central shot from 5 meters should exhibit a high scoring probability. Thus, the closer a shot is to the goal and the better the angle of the shot is, all else being equal, the higher the scoring probability is.

[5] One alternative proxy for scoring chances is the use of ball possession in certain areas of the pitch (Gurpinar-Morgan, 2015). Another alternative is to capture goal-threatening moments, even if the end result is not a shot, and to exclude certain useless shots under defined circumstances (Stratabet, 2017).

Table 1
Factors influencing the scoring probability of a shot.

Location on the pitch That is: – Distance – Angle	Rule setting That is: – Open play – Free kick – Penalty kick	Body part That is: – Foot – Header – Other body parts	Defensive pressure That is (e.g.): – Position of defenders – Position of goalkeeper – Body contact
Motion sequence That is (e.g.): – Out of the air – Out off a dribble – First touch	Player finishing skills That is: – Motor skills – Mental abilities	Goalkeeper skills That is: – Motor skills – Mental abilities	Other That is (e.g.): – Pitch conditions – Spin of the ball – Wind influence

Accurate data on shot locations is readily available and shot location has been used in almost all studies that model scoring probabilities (e.g., Pollard et al., 2004; Rathke, 2017; Caley, 2015).

A second factor that influences the scoring probability of a shot is its rule setting, i.e., whether the shot is taken in open play, from a free kick or from a penalty kick. For example, a shot taken in an open-play situation, where the ball is typically in motion, is different than a shot taken from a free kick, where the ball is at rest and opposing players are required to maintain a certain distance until the ball is touched.[6] Further, the part of the body that is used to shoot also affects the scoring probability. A shot made by foot is generally faster and more precise than a header. Information on the rule setting and the body part used for each shot is widely recorded and is thus typically used for modeling scoring probabilities (e.g., Caley, 2015; IJtsma, 2015; Wright, Atkins, Polman, Jones, & Sargeson, 2011).

An additional factor that is expected to influence the scoring probability of a shot is defensive pressure from the opposing team. If defensive pressure is high, for example, when a defender is right in front of the shot taker or when the defender is already making

[6] According to the rules of the game, goals can also be scored from direct corner kicks and from the kick-off. However, these situations are very rare and are quite similar to free kicks in terms of the rule setting. Therefore, such shots could also be treated as free kicks. By contrast, penalty kicks are a different category in terms of the rule setting because the defending team is much more limited in their options. They are not allowed to position any defender within the penalty box, and the goalkeeper must remain on the goal line until the penalty kick is taken.

body contact with the shot taker, the scoring probability is expected to decrease. Several earlier studies that analyzed video sequences were able to include variables that directly capture defensive pressure. Ensum, Pollard, and Taylor (2004), for example, included the number of outfield players between the shot taker and the goal, as well as a space variable to indicate whether a defensive player is more than 1 meter from the shot taker.

However, this concrete information on the positioning of the defenders at the moment of a shot is not yet systematically available in typical packages offered by professional sport data companies. Thus, most existing models incorporate defensive pressure only indirectly through proxy measures. Caley (2015) includes an indicator variable for shots made after corner kicks because corner kicks are among the most defended actions in football. Further, he includes a variable that indicates whether a shot follows a counter attack, which is a proxy for less defensive pressure. Similarly, IJtsma (2015) includes the game state to model defensive pressure because the defensive pressure of an opponent is expected to increase if the opponent is already leading in the match (and vice versa). Nevertheless, the upcoming availability of tracking data will allow defensive pressure to be calculated more accurately, which will increase the predictive power of estimation models (see e.g., Lucey et al., 2015).

From a more dynamic perspective, the motion sequence of the shot-taking action is another factor that is expected to influence the scoring probability. For example, it matters whether the ball is hit after a fluid run past some defenders or with a volley out of the air after a cross. The motion of the player while shooting differs, which affects the difficulty of the shot. However, such motion sequences are challenging to operationalize empirically compared to the factors discussed above. One would need a video-based or sensor-based scan of body movement, which is not yet available. Therefore, only proxy variables, such as a shot assisted by a cross or a shot preceded by a dribble, are employed to partially account for the motion sequence of the shot taker (see e.g., Caley, 2015; IJtsma, 2015).

The individual finishing skill of the shot taker and the individual skill of the goalkeeper to stop shots on target are two additional factors that are expected to influence the scoring probability of a given shot. Different shot takers have different motor skills that influence the speed and accuracy of the shot and thus the scoring probability. Furthermore, the mental ability of the players, such as dealing with pressure, is also expected to have an impact. Likewise, better goalkeeper skills, such as a short reaction time or an enhanced ability to predict the behavior of the shot taker, decrease the scoring probability of a shot. However, the shot taker's finishing skills and the goalkeeper's shot-stopping skills are difficult to empirically quantify because there are usually insufficient shots observed in the data samples to systematically differentiate between skills and random variations.[7] Therefore, the most current available approaches neglect individual skills and model the scoring probabilities of shots based on the average finishing skills of all shot takers and the average shot-stopping skills of all goalkeepers.

All the factors described thus far are expected to be under the systematic control of the players and teams on the pitch. However, a range of other, more idiosyncratic factors, such as pitch conditions, the spin of the ball and wind, potentially influence the scoring probability of a given shot. These factors are typically not under the systematic control of the players, and their effect on the scoring probability remain mostly diffuse.

Although all these factors are expected to influence the scoring probability of a shot, an effective empirical model should focus on factors that are under the systematic control of the players. Specifically, the model should accurately account for the location of the shot, the rule setting, the body part used, the defensive pressure, the motion sequence, the player's finishing skills and the goalkeeper's skills. However, because some of these factors are difficult to operationalize given the current data availability, only restricted models are empirically feasible at the moment. Nevertheless, the theoretical interpretation of the

[7] By nature, the small-sample issue is more problematic for shot takers than for goalkeepers (because goalkeepers face all the shots from the opposing teams while individual players take only a fraction of the shots of a team) and is even more problematic for defenders than for strikers (because defenders shoot less often than do strikers). Further arguments on the small-sample issue for the identification of shot takers' individual finishing skills can be found in Caley (2015).

estimated scoring probability of a shot is straightforward: how much is a given shot worth in terms of the likelihood that it will lead to a goal based on the specific situation at the time of the shot.

3 Estimation of expected goals

3.1 Data and variables

Our data include shot information on all 7,304 matches played in the English Premier League, the French Ligue 1, the German Bundesliga, the Italian Serie A and the Spanish La Liga in the four seasons from 2013/2014 to 2016/2017.[8] In total, we observe 170,688 shots that resulted in 19,283 goals.[9] For each shot, we have detailed information on the location, the rule setting and the part of the body that was used. In therms of location, we know the exact coordinates of the shot on the football pitch. Following Pollard and Reep (1997), we calculate the shot distance as the Euclidean distance between the shot location and the midpoint between the two goalposts. Furthermore, we calculate the shot angle as the angle between the shot location and the two goalposts, which mimics the player's view of the goal. Hence, the shot angle becomes larger the closer and the more central the location is to the goal, and vice versa.[10] To distinguish between the different rule settings of shots, we create dummy variables for shots made from open play, free kicks, and penalty kicks. For the part of the body, we create dummy variables for shots made with the foot and shots from headers.[11]

Table 2 shows the summary statistics. On average, 11.3% of all shots result in a goal: that is, on average, it takes roughly 9 shots to score one goal. This figure is broadly in line

[8] The dataset is provided by the commercial data provider Gracenote, which has been a Nielsen company since 2017.

[9] Additionally, 594 own goals were scored in these leagues during our sample period. However, own goals are not coded as shots; therefore, they are excluded from the dataset. Moreover, including own goals in our analysis would bias our estimates because they are, by definition, unintended.

[10] Alternatively, Pollard et al. (2004) calculated the angle from a perpendicular line from the nearest goalpost. Following this approach, our results remain qualitatively unchanged.

[11] Shots from other body parts account for only 0.5% of all shots and are thus classified as headers. Classifying shots from other body parts separately does not qualitatively change any of our results.

with the results of Lucey et al. (2015) and Pollard et al. (2004), who report average goal rates of 9.6% and 10%, respectively. The average shot distance for all shots taken is 18.60 meters and ranges from 0.60 to 91.10 meters. The average shot angle is 22.95 degrees, the smallest shot angle is 0.10 degrees, and the largest angle is 170.70 degrees. A total of 93.7% of the shots stem from open play, 5% from free kicks and 1.2% from penalty kicks. Shots struck with the foot account for 84.3% of all shots. The remaining 15.7% of the shots are headers.

Table 2
Summary statistics for shots.

Variable	Mean	Std. dev.	Min.	Max.
Goal	0.113	0.317	0	1
Location				
Distance	18.60	7.43	0.60	91.10
Angle	22.95	12.81	0.10	170.70
Rule setting				
Open play	0.938	0.243	0	1
Free kick	0.050	0.219	0	1
Penalty kick	0.012	0.110	0	1
Body part				
Foot	0.843	0.364	0	1
Header	0.157	0.364	0	1

Notes: The number of observation is 170,688. The shot distance is measured in meters and the shot angle is measured in degrees.

3.2 Empirical model and estimation

Following Pollard et al. (2004) and Wright et al. (2011), we employ a logistic regression analysis to estimate the probability of a goal, which is our binary response variable. Formally, our model takes the form

$$\begin{aligned}\mathrm{Ln}\left[\frac{P(Goal_i=1)}{P(Goal_i=0)}\right] &= \beta_0 + \beta_1\, Distance_i + \beta_2\, Angle_i + \beta_3\, Free\ kick_i \\ &\quad + \beta_4\, Penalty\ kick_i + \beta_5\, Header_i + \varepsilon_i,\end{aligned} \tag{1}$$

where the subscript i denotes a shot. The base category for the rule setting is open play and the base category for the body part is a shot made by foot.[12]

Table 3 shows the estimation results for Equation 1. For example, the coefficient for the shot distance is -0.1307, indicating that for every meter from the goal, the odds of scoring decrease by 12.3%.[13] Furthermore, the odds of scoring increase for each additional degree of the shot angle. Free kicks and penalty kicks have a higher scoring probability than do ordinary shots from open play, whereas headers have a low scoring probability compared to shots made with the foot.

Table 3
Estimation results from logistic regression.

	Goal (0/1)
Intercept	-0.696***
	(0.065)
Distance	-0.1307***
	(0.003)
Angle	0.029***
	(0.001)
Free kick	1.049***
	(0.050)
Penalty kick	2.193***
	(0.052)
Header	-1.054***
	(0.023)
Pseudo R^2	0.191
N	170,688

Notes: Logit estimates for Equation 1 are displayed. The binary dependent variable indicates whether a goal resulted from a shot (1) or not (0). Robust standard errors are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Based on these coefficients, we predict the scoring probability of each shot in our sample. For example, an open-play long-range shot by foot from 30 meters with an angle of 11 degrees has a predicted scoring probability of approximately 1%. By contrast, an open-play shot by foot from 5 meters with an angle of 70 degrees has a predicted scoring probability of 66%.[14]

To derive the number of expected goals in a match, we aggregate the estimated scoring probabilities for all the shots taken by each team. To illustrate this process, Table 4 shows

[12] Alternatively, the probabilities of free kicks and penalty kicks could be estimated in separate models to account for the different relationships between the likelihood of scoring a goal and the distance and angle from the goal (see e.g., Caley, 2015). However. we refrain from this approach to keep our estimation as simple as possible.

[13] The odds ratio is equal to $e^{-0.1307} = 0.877$.

[14] Calculation: $-0.696 - 0.1307 * 5 + 0.029 * 70 = 0.681; e^{0.681} = 1.98; 1.98/(1 + 1.98) \sim 66\%$.

all the shots made during the match between Arsenal and Manchester United on May 7, 2017. For example, in the 65th minute of the game, Manchester United's player Wayne Rooney took a shot from a direct free kick 26.8 meters from the goal for which our model predicts a scoring probability of approximately 6%. In total, the 8 shots made by Arsenal add up to chances worth 0.97 goals, and the 12 shots made by Manchester United add up to chances worth 0.90 goals. Thus, even though Manchester United had four more shots than did Arsenal, both teams produced chances of almost equal value in the match. Nevertheless, Arsenal won the game 2–0, as Granit Xhaka and Danny Welbeck scored from their shots in the 54th and 57th minutes. Notably, the goal made by Xhaka resulted from a long-range shot that was 32 meters from the goal, for which our model predicts a scoring probability of approximately 1%. However, the shot was deflected from the back of a Manchester United midfielder and luckily found its way over the goalkeeper into the net.

Table 4
Example calculation of the scoring probabilities of all shots made during a match played between Arsenal and Manchester United.

Minute	Player name	Distance	Angle	Rule setting	Body part	Scoring probability of the shot	
						Arsenal	ManU
2	Wayne Rooney	14.3	24.1	Open play	Header		0.05
5	Anthony Martial	12.8	19.7	Open play	Foot		0.14
9	Aaron Ramsey	14.4	17.3	Open play	Foot	0.11	
25	Wayne Rooney	9.3	42.6	Open play	Header		0.15
26	Danny Welbeck	10.7	32.6	Open play	Foot	0.24	
30	Danny Welbeck	12.5	27.9	Open play	Foot	0.18	
31	Alex Oxlade-C.	25.1	16	Open play	Foot	0.03	
32	Wayne Rooney	11.3	31.1	Open play	Foot		0.22
54	Granit Xhaka	31.7	12.7	Open play	Foot	0.01	
57	Danny Welbeck	5.9	62.2	Open play	Header	0.33	
65	Wayne Rooney	26.8	13.2	Free kick	Foot		0.06
66	Juan Mata	26.6	15.2	Open play	Foot		0.02
68	Granit Xhaka	25.1	16.6	Open play	Foot	0.03	
74	Wayne Rooney	26.9	15.5	Open play	Foot		0.02
78	Anthony Martial	29.3	14	Open play	Foot		0.02
81	Wayne Rooney	28.9	14.3	Open play	Foot		0.02
87	Alexis Sanchez	22.2	15.8	Open play	Foot	0.04	
89	Wayne Rooney	16	14	Open play	Foot		0.08
91	Marcus Rashford	19.6	16.7	Open play	Foot		0.06
92	Scott McTominay	20.4	18.9	Open play	Foot		0.06
Sum of the scoring probabilities for each team						0.97	0.90

Notes: The table shows all shots and their estimated scoring probabilities from the match between Arsenal and Manchester United played on May 7, 2017. The sum of all scoring probabilities from one team corresponds to the expected goals of the respective team.

Note that our estimated scoring probabilities indicate a comparable likelihood based on the shot characteristics considered by our expected goal model, i.e., the location of the shot, the rule setting of the shot and the body part used for the shot. The true likelihood that a shot will become a goal can still deviate substantially from the estimated likelihood in a particular case if the shot differs in a characteristic that is not considered in this model. For example, if a shot is taken after dribbling past a goalkeeper, the scoring probability will be underestimated in our model because we do not account for the empty net situation. Nevertheless, the main objective of quantified scoring chances as a performance evaluation tool is to make chances comparable in a systematic and objective way, with full knowledge that the estimated and true scoring probability can deviate for a particular chance.

4 Comparison of performance evaluation measures

To be useful for decision making inside football clubs, a performance evaluation measure based on expected goals must contain more relevant information about a team's true performance on the pitch than do match outcomes. To this end, the consistency of the measure is crucial because true performance on the pitch is what clubs want to develop over time through a combination of squad quality, coach quality and execution quality of strategies and tactics developed off the pitch. To capture such developments within the course of a single season and from season to season in a consistent way, the influence of random variation on any measure of performance must be low.

Expected goals consider all the scoring chances, whereas match outcomes are based solely on the few actual goals. Thus, expected goals should be less prone to the randomness associated with match outcomes because they are based on a much larger number of actions that represent a team's true performance on the pitch.[15] However, the informativeness of expected goals in terms of the true performance on the pitch will also be impaired if

[15] In an average match in our sample, the number of goals is 2.7 and the number of shots is 23.4.

some of the qualities of a team are not systematically captured by the model. Ultimately, the problem comes down to an empirical question of whether one of the two effects is dominating the other while capturing the true performance as accurately as possible.

To address this empirical question, we compare the informativeness of both performance evaluation measures by testing how well they predict the sporting success a team will have in the future. Specifically, we test how well the number of points won in previous matches and the difference between expected goals created and allowed in the same matches predict the number of points won by a team in the following matches. In the first step, for any given matchweek, for every team in our sample, we estimate a univariate linear regression model in which the dependent variable is the number of points won in the following ten matches, and the independent variable is either the number of points won in the previous ten matches or the difference between the expected goals created and allowed in the previous ten matches. In the analysis, we follow Caley (2015) and employ a rolling perspective across seasons.[16]

Table 5
Estimation results from the ordinary least squares regression for ten matches.

	Number of points won next ten matches	
	(1)	(2)
Number of points won in previous ten matches	0.506***	
	(0.008)	
Difference between expected goals created and allowed in previous ten matches		0.546***
		(0.007)
R^2	0.253	0.320
N	12,119	12,119

Notes: Robust standard errors are given in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

To evaluate the predictive accuracy of the two metrics, we use the goodness-of-fit measure R^2, which represents the percentage of the variance in future success (i.e., points won) that is explained by each metric from past matches in our model. Column (1) of Table 5 shows that the number of points won in the previous ten matches explains 25.3%

[16] To ensure that all the information about the ten previous matches and the ten following matches is included, we cannot consider the first ten matchweeks of the 2013/2014 season and the last nine matchweeks of the 2016/2017 season because we lack data from the 2012/2013 and 2017/2018 seasons that are required to create the respective measures. Similarly, we exclude the matchweeks at the beginning of a season for teams that were promoted during our sample period and the matchweeks at the end of a season for teams that were relegated during our sample period because we lack data from the 2nd divisions.

of the variation in the number of points won in the following ten matches. By contrast, Column (2) of Table 5 shows that the difference between the expected goals created and allowed in the previous ten matches explains 32.0% of the variation in the number of points won in the following ten matches. These results indicate that expected goals contain more relevant information about a team's true performance on the pitch in recent matches.

In the second step, we run the linear regression 1,444 times for every combination from 1 to 38 previous matches and from 1 to 38 following matches to test whether the previous finding is robust to using different numbers of past and future matches.[17] Figure 1 displays all resulting R^2 values. The light gray shading indicates the R^2 values from the regressions on the number of points won, and the dark gray shading depicts the R^2 values from the regressions on the difference between the expected goals created and allowed. For every combination of the number of previous matches and the number of following matches in Figure 1, the R^2 calculated using expected goals is higher than the R^2 calculated using the number of points. Thus, expected goals contain more information about true performance on the pitch for any sequence of matches within the horizon of a full season.

These results show that the advantage of expected goals, achieved through successfully filtering out the random components of actual goals, outweighs the disadvantage of a potential informational loss related to team qualities that are not captured by our model. This results is noteworthy since our estimation of shots' scoring probabilities follows a very basic model and accounts for only three of the factors outlined in Section 2. A more comprehensive model that accounts for additional factors should cause the use of expected goals from previous matches to have an even greater informational advantage over using match outcomes from previous matches.

[17] Because we must omit matchweeks depending on the numbers of previous matches and the number of following matches, the number of observations varies for the 1,444 regressions. The lower bound is 5,812 matchweek-team observations in the regression in which we use information from the past 38 matches to predict the success in the next 38 matches. The upper bound is 14,477 matchweek-team observations in the regression, where we use information from only the last match to predict the success in the next match.

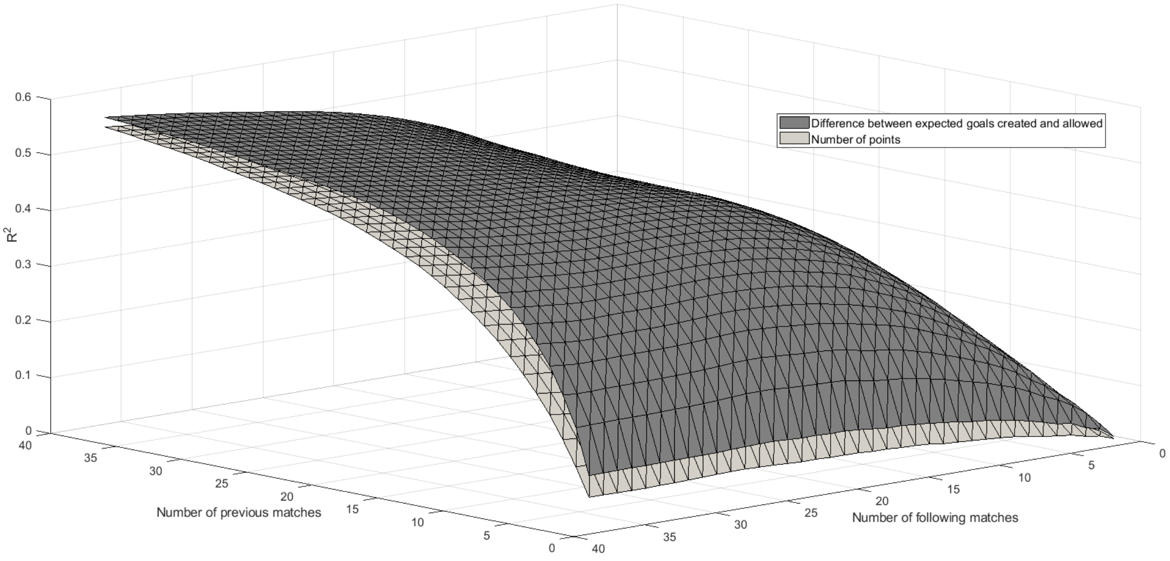

Figure 1
R^2 values from the regressions of points won in following matches on performance in previous matches for different numbers of previous and following matches. The dark and light gray shading represent the R^2 values when the performance is measured by the difference between expected goals created and allowed and by points won, respectively.

5 Overperformance and underperformance of expected goals

In the previous section, we have shown the informational advantage of expected goals over match outcomes from a full sample perspective. However, to interpret the information of expected goals for a particular club, we must address differences between expected goals and actual goals (i.e., match outcomes) from the perspective of an individual case. In particular, substantial deviation in a team's actual goals and expected goals over a given time period could be due to either the randomness of actual goals or qualities of a team that are not captured in the model. For example, a team that has the ability to create scoring chances with a lower level of defensive pressure might score more actual goals in a particular situation than do other teams. However, because our model does not account for differences in defensive pressure, such a team will systematically overperform the expected goal estimates. By contrast, a team with lower than average abilities to create chances that face less defensive pressure might systematically score fewer actual goals than expected. While such an overperformance or underperformance is systematic, an overperformance or underperformance due to randomness is unsustainable. To improve the decision making by managing the randomness inherent to match outcomes, we are interested in only the latter type. Thus, we must assess to what extent the overperformance or underperformance of expected goals is systematic.

A precise separation of the influence of random forces and the influence of existing qualities that are not considered in the expected goal estimates is hard to achieve because a team's true underlying quality is not time-constant due to, for example, player transfers or coach changes. Nevertheless, we can identify the teams with the highest and lowest long-term overperformance and underperformance within the scope of our sample to construct some feasible thresholds. In particular, we calculate the overperformance and underperformance as the ratio between all actual goals scored (conceded) and all expected goals created (allowed) of a team within the 152 matches played between the 2013/14 and 2016/17 seasons.[18] The higher the offensive ratio, the more goals a team scored compared to expected goals created and the higher the defensive ratio, the more actual goals a team conceded compared to the expected goals allowed.

Panel A in Table 6 shows that Real Madrid, Napoli, Borussia Mönchengladbach and FC Barcelona are the four teams with the highest offensive ratios. For example, Real Madrid scored 24.2% more actual goals than expected during the four seasons in our sample. It appears to be reasonable that these four teams, which are all known for exceptional offenses in terms of either their individual star players and/or their coaches' tactical execution[19], form an upper bound of systematic offensive overperformance. The teams with the lowest offensive ratios are Crystal Palace, Sunderland, FC Nantes and West Bromwich Albion. Those teams underperformed considerably, scoring approximately 10% fewer goals than

[18] We exclude newly relegated or promoted teams and calculate the ratios only for teams that were consecutively in our sample.

[19] In two of the four seasons included in our sample, Borussia Mönchengladbach was coached by Lucien Favre, who is especially known for an offensive style that outperforms expected goals (Raman, 2017). Similarly, Napoli has been coached by Maurizio Sarri in two of the four seasons in our sample, a coach who also has a strong reputation for an offensive tactical execution that outperforms expected goals (Kwiatkowski, 2018).

expected, which could be explained by the fact that these teams are systematically below average in some team qualities that are not fully considered in our model. Panel B in Table 6 shows the teams with the best and worst defensive ratios. Juventus, Atlético Madrid, FC Bayern München and Manchester United, teams that are known for their exceptional defensive strength and their exceptional goalkeepers (i.e., Gianluigi Buffon, Jan Oblak, Manuel Neuer and David de Gea), conceded between approximately 28% and 19% fewer goals than estimated by our model. By contrast, Werder Bremen, FC Lorient, RCD Espanyol and Toulouse FC conceded approximately 20% more goals than expected.

If a team exhibits a ratio considerably above or below these thresholds during a shorter evaluation period, it seems very unlikely that the ratio is systematically driven by underlying qualities and much more likely that the difference is unsustainable due to randomness in actual goals scored or goals conceded. Accordingly, the thresholds can provide insights to the decision makers of clubs in order to evaluate situations in which actual goals and expected goals diverge over a given sample of matches.[20]

To analyze how often teams temporarily perform above or below the long-term thresholds presented in Table 6, we calculate the offensive and defensive ratios based on a short sequence of five matches, i.e., matchweeks 1-5, 6-10, 11-15, 16-20, 21-25, 26-30, and 31-35. Table 7 displays the 25^{th} percentile (Q_1), the 50^{th} percentile (Q_2) and the 75^{th} percentile (Q_3) of the ratio distribution for these sequences of five matchweeks. Because the values of Q_1 (Q_3) are mostly smaller (larger) than the thresholds, in more than 25% of the observations a team scored temporarily below (above) the thresholds. Thus, unsustainable overperformance or underperformance is observed very frequently if we look at sequences of a small number of matches.[21] To note a particular extreme case, FC Villarreal created 37 shots worth 4.3 expected goals and allowed 57 shots worth 7.2 expected goals to their

[20] The thresholds should become narrower if more relevant factors are considered in the expected goal model. For example, if we include the current score of the match at the moment of the shot in our model to proxy defensive pressure to some extent (see asdf), the highest offensive ratio we observe for a team over the full sample period decreases from 1.242 to 1.206. Similarly, the lowest offensive ratio increases from 0.885 to 0.901).

[21] Even if we extend the match sequences, the number of observations in which teams perform above or below the thresholds remains considerable. For example, for a sequence of 10 matches, 21% (32%) are above (below) the offensive ratio and 14% (27%) are below (above) the defensive ratio.

Table 6
Offensive and defensive over- and underperformance of expected goals.

Panel A: Offensive overperformance and underperformance

Team	Expected goals created	Actual goals scored	Offensive ratio	Ratio rank
Real Madrid	352.6	438	1.242	1
Napoli	262.6	321	1.223	2
Borussia Mönchengladbach	183.4	224	1.222	3
FC Barcelona	358.8	438	1.221	4
.				
.				
.				
West Bromwich Albion	174.2	158	0.907	65
FC Nantes	158.9	142	0.894	66
Sunderland	167.3	149	0.891	67
Crystal Palace	191.0	169	0.885	68

Panel B: Defensive over- and underperformance

Team	Expected goals allowed	Actual goals conceded	Defensive ratio	Ratio rank
Juventus	130.2	94	0.722	1
Atlético Madrid	138.2	100	0.724	2
FC Bayern München	108.2	80	0.740	3
Manchester United	177.6	144	0.812	4
.				
.				
.				
Toulouse FC	183.9	213	1.159	65
RCD Espanyol	187.6	226	1.204	66
FC Lorient	191.4	231	1.207	67
Werder Bremen	215.1	260	1.209	68

Notes: Offensive and defensive over- and underperformance is calculated for all 68 teams that were consecutively in the sample between season 2013/14 and 2016/17 (152 matches). The ratios are calculated as actual goals divided by expected goals. The table only displays the four highest- and lowest-ranked teams.

opponents between matchweeks 16 and 20 in the 2015/16 season. However, they scored 7 goals (offensive ratio of 1.63) and conceded only one goal (defensive ratio of 0.14) during these five matches. Accordingly, even though that FC Villarreal was not very dangerous in terms of chance creation or defensively very solid in terms of chance allowance, the team gained an almost perfect 13 points out of the five matches based on the numbers of actual goals scored and conceded due to their overperformance.

Table 7
Quartiles of offensive and defensive short-term ratios.

Overperformance and underperformance	N	Q_1	Q_2	Q_3
Offensive ratio	2,672	0.760	1.002	1.289
Defensive ratio	2,672	0.758	1.003	1.293

Notes: The ratios are calculated as actual goals divided by expected goals over sequences of five matches. Q_1, Q_2, Q_3 refer to the 25^{th}, 50^{th}, and 75^{th} percentiles, respectively.

Overall, there is a considerable number of situations for clubs in which the discrepancy between actual goals and expected goals cannot be explained by systematic overperformance or underperformance. Rather, unsustainable overperformance or underperformance due to randomness appears to be the driver of ratios that lie beyond the thresholds of systematic overperformance or underperformance. In the next section, we derive a simple tool to identify critical situations in which randomness is likely to disguise true performance on the pitch.

6 Expected goals and decision making

In situations where match outcomes indeed reflect the true performance on the pitch, the potential for misjudgment is limited. However, in situations where match outcomes are influenced by numerous random components and misrepresent the true performance on the pitch over a series of matches, misjudgments can arise. One way to identify such situations is to compare the ranking of a team in the official league table to a ranking based on the difference between expected goals created and allowed by a team. As an example, Table 8 shows the official league table and the ranking based on expected goals for the English Premier League halfway through the 2016/17 season. At that time, Chelsea was ranked first in the official league table with 49 points. By contrast, Manchester City had the highest ranking based on expected goals, with generated chances worth 42.1 goals, allowed chances worth 19.2 goals, and a positive difference of +22.8.

For 15 of the 20 teams, the rank difference is within three and is thus relatively small. However, there are also teams that are substantially over- or under-rewarded in the official league table. On one hand, Burnley ranks 11th in the official league table while ranking 20th, and thus last, based on expected goals. On the other hand, Leicester City ranks 15th in the official league table even though their performance based on expected goals would rank them much better in the 9th position.

Table 8
Assessment of team performance in the English Premier League halfway through the 2016/17 season.

Official league table			Ranking based on expected goals				
Rank	Club	**Points**	Rank	Club	created	allowed	Δ
1	Chelsea	**49**	1 (+4)	Manchester City	42.1	19.2	**+22.8**
2	Liverpool	**43**	2 (0)	Liverpool	39.7	17.4	**+22.3**
3	Arsenal	**40**	3 (+1)	Tottenham Hotspur	37.0	18.9	**+18.2**
4	Tottenham Hotspur	**39**	4 (+2)	Manchester United	34.7	18.0	**+16.7**
5	Manchester City	**39**	5 (–4)	Chelsea	32.6	16.5	**+16.1**
6	Manchester United	**36**	6 (–3)	Arsenal	34.6	20.3	**+14.3**
7	Everton	**27**	7 (+2)	Southampton	30.5	21.7	**+8.7**
8	West Bromwich A.	**26**	8 (–1)	Everton	25.0	22.8	**+2.2**
9	Southampton	**24**	9 (+6)	Leicester City	26.2	26.2	**–0.1**
10	Bournemouth	**24**	10 (+2)	West Ham United	27.0	30.6	**–3.6**
11	Burnley	**23**	11 (–1)	Bournemouth	24.9	28.6	**–3.7**
12	West Ham United	**22**	12 (–4)	West Bromwich A.	20.7	25.4	**–4.7**
13	Watford	**22**	13 (+1)	Stoke City	21.8	28.5	**–6.7**
14	Stoke City	**21**	14 (+3)	Crystal Palace	26.2	33.2	**–6.9**
15	Leicester City	**20**	15 (–2)	Watford	19.1	29.1	**–10.0**
16	Middlesbrough	**18**	16 (0)	Middlesbrough	16.8	30.0	**–13.2**
17	Crystal Palace	**16**	17 (+3)	Swansea City	25.3	40.2	**–15.0**
18	Sunderland	**14**	18 (0)	Sunderland	21.0	37.6	**–16.6**
19	Hull City	**13**	19 (0)	Hull City	18.6	38.9	**–20.2**
20	Swansea City	**12**	20 (–9)	Burnley	17.2	38.0	**–20.8**

Notes: Both tables indicate the ranking of the teams after matchweek 19 (out of a total of 38 matchweeks). The brackets in the rank column for the ranking based on expected goals refer to a team's rank difference against the official league table. Δ indicates the difference between expected goals created and allowed.

To identify and illustrate situations with significant discrepancies based on the data reported in Table 8 in a non-technical way, the graph in Figure 2 plots each team's ranking in the official league table on the y-axis and its ranking based on expected goals on the x-axis. The identity line marks where a team would have the same ranking in both tables. Teams located below the identity line have a better rank in the ranking based on expected goals compared to the official league table and, vice versa.

A quick glance at Figure 2 clearly shows that Burnley is over-rewarded and Leicester City is under-rewarded in the official league table halfway through the 2016/17 season. In general, the further away from the identity line a team is located, the more likely it is that randomness played an important role in the team's results. The larger the difference between the rankings of the two tables is, the more severe it is when decision makers judge the performance of a team based only on the official league table. Thus, any outlier should trigger an alert and make decision makers aware that the rank in the official league table might be substantially driven by randomness. As we discussed in Section 5, it

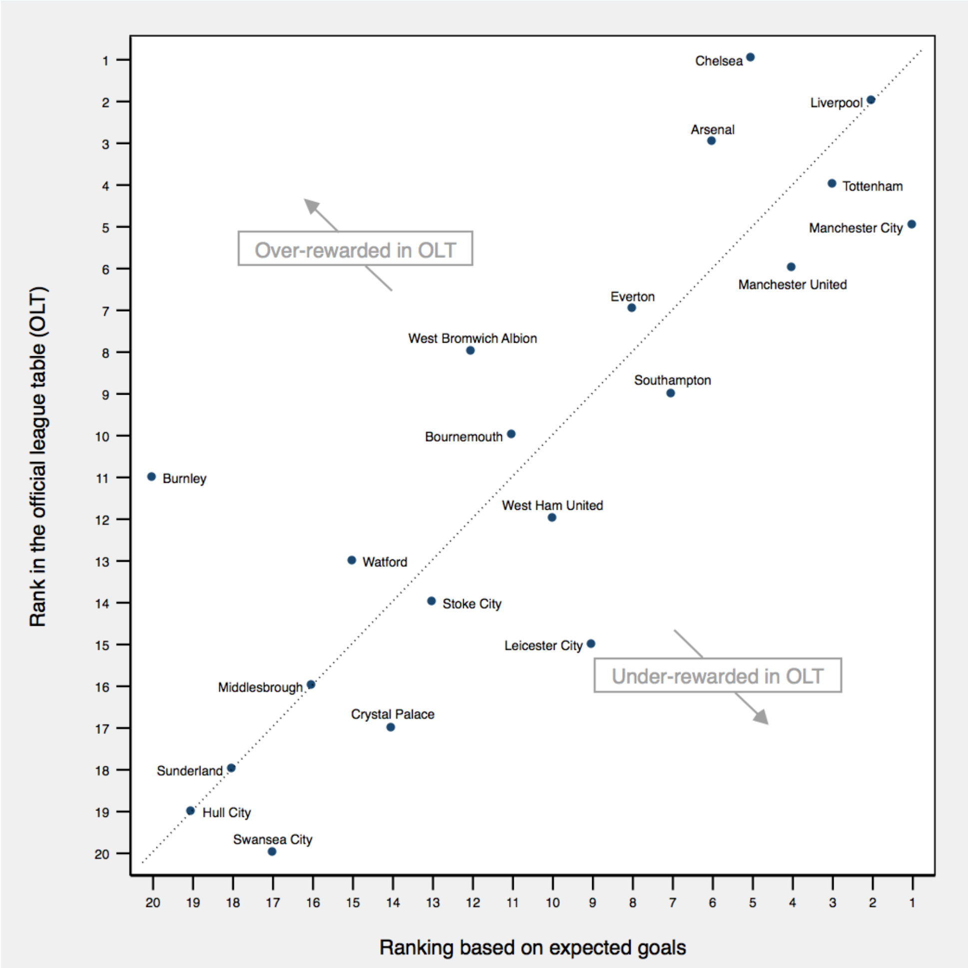

Figure 2
Rank in the official league table versus the ranking based on expected goals for the English Premier League teams halfway through the 2016/17 season. The ranking based on expected goals sorts the teams by their difference between expected goals created and allowed.

remains possible that some discrepancies are due to team quality characteristics that the expected goal model does not account for. Nevertheless, any alert can be further examined to judge whether there are legitimate reasons to conclude that the overperformance or underperformance is systematic.

To investigate the discrepancies between the official league table and the ranking based on expected goals from a broader perspective, we show all team observations in our sample halfway through each season in Figure 3. Specifically, the figure plots the rankings based on expected goals against the official league ranking for all 392 team-season rankings after matchweek 19 of the English Premier League, the French Ligue 1, the Italian Serie

A and the Spanish La Liga and after matchweek 17 of the German Bundesliga for the four seasons from 2013/14 to 2016/17. Because many of these team-season observations overlap, the size of the bubbles represents the number of observations for a specific rank combination. The largest bubble is located in the upper-right corner because the most common combination is that a team is ranked 1st in both the official league table and based on expected goals. This result appears to be reasonable because the ranking is bounded at the 1st rank and the very dominant teams usually also perform very well in terms of expected goals. Overall, the results show that, even though the majority of the teams are located quite close to the identity line, large discrepancies of 7 or more ranks occur for a substantial 11% of all cases at the halfway point of a season. These cases can be found in the upper-left and the lower-right corners and are the situations most prone to misjudgments.

In the following, we discuss three examples where teams were ranked much worse in the official league table than their ranking calculated using their performance based on expected goals, i.e., the lower-right corner in Figure 3. In other words, these cases include teams that generally performed well in creating chances and in preventing opponents from doing so but did not translate this accomplishment into positive match results and a corresponding ranking in the league table. First, VfB Stuttgart were ranked 16th in the official league table but a lofty 6th based on expected goals after the first half of the 2015/16 season. Nevertheless, the club officials decided to dismiss the coach Alex Zorniger after matchweek 13 (VfB Stuttgart, 2015). Second, the Italian Serie A club Cagliari Calcio fired its coach Zdenek Zeman after matchweek 16 in the 2014/15 season, when it was ranked 18th in the official league table (The Guardian, 2014). In the ranking based on expected goals, however, Cagliari was ranked 6th which suggests a much better underlying performance. Third, Real Betis dismissed Pepe Mel after matchweek 15 in the 2013/2014 season when they were placed last in the official league table, but they were in a satisfactory 8th place in the ranking based on expected goals.

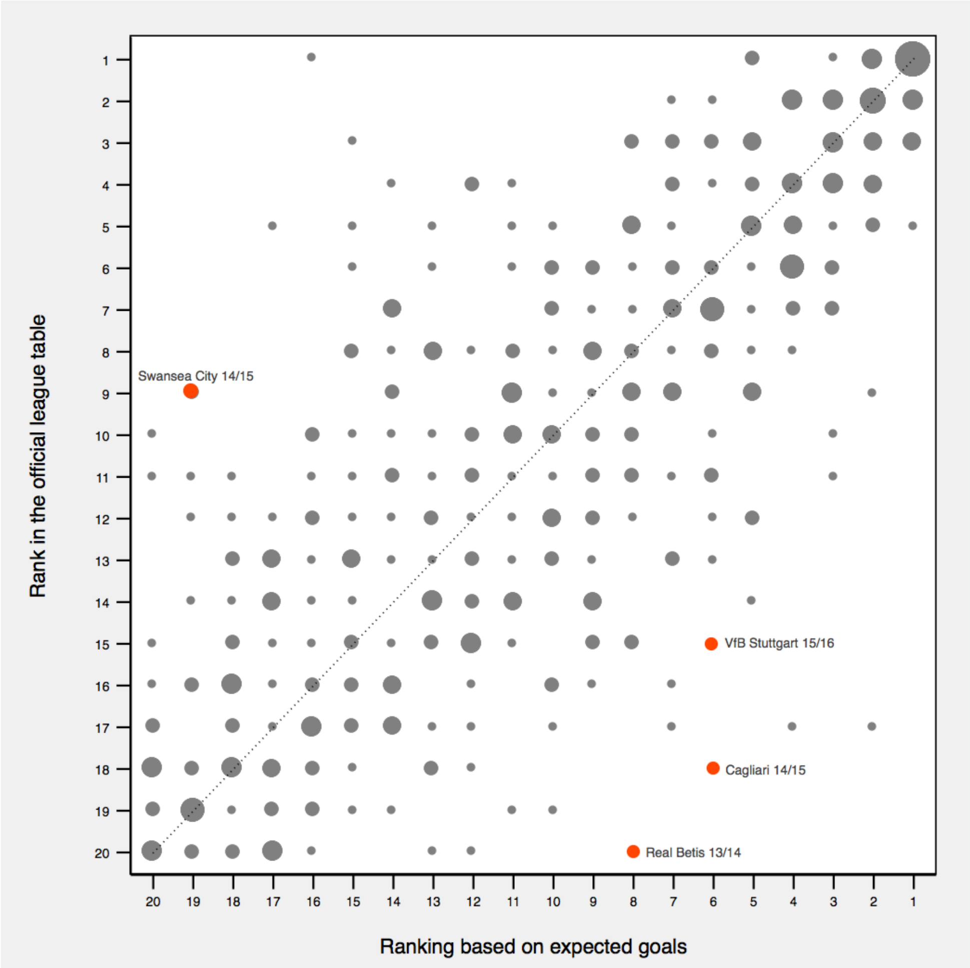

Figure 3
Rank in the official league table versus the ranking based on expected goals for all teams in the English Premier League, the French Ligue 1, the Italian Serie A, the Spanish La Liga and the German Bundesliga at the halfway point of each season from 2013/14 to 2016/17. The ranking based on expected goals sorts the teams by the difference between expected goals created and allowed.

All the clubs in these three examples took action after being on a disappointing rank in the official league table, although their performance based on expected goals ranked them at least ten ranks higher. Furthermore, the offensive and defensive ratios indicate that their underperformance of expected goals was likely to be unsustainable. For example, the offensive ratio of Real Betis calculated from the beginning of the season up to the dismissal decision was 0.734 and the defensive ratio was 1.632, implying that they scored approximately 27% less goals and conceded approximately 63% more goals than expected based on the scoring chances they created and allowed.[22] Thus, Real Betis was

[22] The offensive (defensive) ratios for Cagliari were 0.830 (1.442) and for VfB Stuttgart 0.719 (1.304), respectively.

substantially under-rewarded in terms of goals actually scored out of the chances they created and in terms of goals they actually conceded out of the chances they allowed to their opponents.

Unfortunately, non of these three clubs were able to significantly rebound in the official league table after replacing their coaches, and all were relegated to the 2nd division at the end of the season. Interestingly, not only were the teams not able to improve their match outcomes and their rankings in the official league table, but their rankings based on expected goals also declined during the rest of the season. This suggests that their true performance on the pitch was worse under the new coach than it was under the coach who was replaced. Although we do not know exactly what would have occurred if the old coach had been allowed to continue, the discrepancy between the two table ranks at the moment the decision makers took action and the subsequent development in the remainder of the season provides suggestive evidence that replacing the coach sealed the team's fate and had costly consequences for the club.

For situations where teams are located in the upper-left corner of Figure 3, the perception of a team's true performance on the pitch might be overly optimistic, which can also lead to misjudgments and flawed decision making. For example, Swansea City was ranked 9th in the official league table halfway trough the Premier League 2014/15 season, while being ranked only 19th based on expected goals. The defensive ratio of 0.718 upon this point in time suggests that Swansea conceded approximately 28% fewer goals than expected, a value that cannot be explained by systematic overperformance. The offensive ratio of 1.157 was less pronounced but still indicates an offensive overperformance of approximately 15%.

The discrepancy between the table ranks persisted until the end of the season when Swansea finished 8th in the official league table – the highest finish in the club's history – while it was ranked 20th, and thus last based on expected goals. After the club-record finish in the official league table, the decision makers at Swansea decided to give its young coach Garry Monk, who had taken over a few months before the beginning of the 2014/15 season, a contract extension until 2018, which was accompanied by a significant salary increase (Talksport, 2015). At that point in time, the club's chairman Huw Jenkins described the contract extension as a "[...] deserved reward for the fantastic season we've just had [...]." (BBC, 2015, para. 6).

However, the excitement about Garry Monk's work in the 2014/15 season at Swansea did not last long. Although Swansea had a good start in the first four matchweeks of the 2015/16 season, they dropped heavily between matchweeks 5 and 15. After this stretch of ten matches with seven losses, two draws and only one win, Swansea was ranked 15th in the official league table. At that point, Swansea took action again and dismissed Garry Monk. Chairman Jenkins now stated, "To find ourselves in our current situation from where we were in the first week of September, and considering the drop of performance levels and run of results over the last three months, it has brought us to this unfortunate decision today." (The Guardian, 2015a, para. 4). This decision became much more costly due to the extension of Monk's contract until 2018 just a few months earlier.

In defense of the dismissal decision, and probably also in defense of the contract extension a few months earlier, Jenkins further argued that "[...] when you take into account the excellent campaign we had last season when we broke all club records in the Premier League, nobody foresaw the position we would be in at this moment in time." (The Guardian, 2015a, para. 5). We respectfully disagree. Based on the ranking of expected goals, warning signs were clearly evident at the moment Swansea made the decision to extend the contract of Garry Monk. Swansea was ranked more than 10 ranks worse in the ranking based on expected goals with an offensive overperformance of approximately 10% more goals scored than expected and a defensive overperformance of approximately 20% less goals conceded than expected. Given this weak performance of Swansea in terms of chance creation and chance prevention during the 2014/15 season, it seemed foreseeable that the team would likely not be able to come close to repeating the results they achieved in the previous season. Thus, we argue that the decision makers of Swansea did not properly judge the team's true performance on the pitch in the 2014/15 season based on the official league table. Instead, they might have been misled by the overly positive match outcomes and made a hasty and costly contract-extension decision for the club.

Overall, we suggest that expert decision makers inside clubs can make mistakes when forming judgments and making decisions in situations in which random forces have highly impacted the outcomes. The rank-comparison chart is a simple tool that can be used to create a new awareness of situations that are sensitive to flawed judgment and decision making. Considering the ever-increasing stakes of decisions made in European club football, we expect clubs to show a growing need for such approaches that complement the existing practice of outcome-based performance evaluation to improve the overall quality of their decision making.

7 Conclusion

In this paper, we contribute to the improvement of performance evaluation and decision making in European club football in several ways. First, we introduce expected goals based on quantified scoring chances as an alternative method to forming a judgment about team performance in football. We provide evidence that expected goals are a superior source of information to match outcomes by showing that the difference in a team's expected goals created and those allowed in previous matches is a better indicator of its subsequent results than are the number of points the team won in those matches. Therefore, we propose that considering expected goals will generally allow for a more objective assessment of a team's true performance on the pitch than would considering actual match outcomes. Indeed, our analysis at the individual club level suggests that an overperformance or underperformance of expected goals in the short-run is often due to randomness and thus unsustainable.

Second, we illustrate how this method, which is readily available, can be applied to identify situations for which the decision makers of clubs are prone to make misjudgments.

By plotting the teams' rankings in the official league table against their rankings based on expected goals, situations where a team's results are much better or worse than their expected goals would imply become instantly visible. Using this information, clubs' decision makers should be able to avoid the fallacy of inferring poor performance from poor match outcomes or inferring good performance from good match outcomes in situations where this link is not present. The costs of applying such a method appear to be minimal compared to the enormous costs of poor decision making. Thus, the method should also be economically viable.

Third, we lay the groundwork for the development of a new mindset in professional football clubs. Thus far, clubs have tended to underestimate the large role that randomness plays in football results. Instead, common practice is to consider match outcomes as the most important indicator of a team's true performance on the pitch and, ultimately, for the quality of work of the club's sporting personnel. We recommend that clubs begin to incorporate more process-oriented and data-driven evaluation strategies into their decision making processes. Certainly, a range of early-adapting clubs have already moved in this direction at a fast pace (see e.g., The Guardian, 2015b; NY Times, 2017). However, from an industry-wide perspective, this trend seems to remain in its infancy. Against this background, the concept of expected goals as a complementary information source for performance evaluation can be seen as a starting point for a new way of thinking in boardrooms and on other levels of decision making in European football clubs.[23]

Our study is subject to at least two limitations. First, there might be more clubs that already incorporate analytical, data-driven approaches into their decision making processes than is currently known. In fact, it is reasonable to believe that some of the progress inside the clubs remains hidden because clubs take care not to disclose their

[23] One may ask why European club football has thus far been strongly resistant to such data-analytic approaches. One reason is that detailed match data were long presented without context and meaningful analysis; therefore, clubs were skeptical that anything useful could be learned (Walerius, 2017). As described by football writer Gabrielle Marcotti "I think when ... data first became available there was a lot of what I consider bad data or meaningless decontextualized data, like, you know, distance covered or passing percentage or possession percentage and I think a lot of the managers looked at this and quite clearly, quite soon realized that this is kind of nonsense on its own." (Walerius, 2017, para. 33). Thus, the lack of prospects for the meaningful use of these data has driven clubs toward inaction (Anderson & Sally, 2014).

specific approaches and treat this information as proprietary. However, the overall mindset that can be inferred from the clubs' actions and external representation does not make a strong case for an alternative inner view. Second, the rankings based on expected goals are calculated as the difference between expected goals created and expected goals allowed by a team. This could bias the ranking if many of the expected goals that are either created or allowed stem from just one or a few matches. Alternatively, we could derive a team's expected number of points based on the estimated scoring probabilities of the shots via statistical simulation. The appealing element of this approach is that every match receives the same weighting, which is comparable to the construction of the official league table. However, a simulation of expected points is much more difficult to grasp than the ranking based on expected goals because the latter ultimately translates a football-intuitive question into numbers: How many chances did we create and how many chances did we concede compared to our opponents?

In any case, the raison d'être of the idea presented in this paper is not that our ranking based on expected goals should represent the true performance on the pitch in a perfect way. Instead, the key message is that such a ranking based on expected goals can act as a highly valuable warning system for decision makers to mitigate the tendency of overlooking the influence of randomness in match outcomes and to improve judgment of team performance.

References

Anderson, C., & Sally, D. (2014). *The numbers game. Why everything you know about football is wrong.* Penguin Books.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of personality and social psychology*, *54*(4), 569.

BBC. (2015, July). *Swansea City: Boss Garry Monk signs new three-year deal.* Retrieved from http://www.bbc.com/sport/football/33474191

Caley, M. (2015, October). *Premier League projections and new expected goals.* Retrieved from https://cartilagefreecaptain.sbnation.com/2015/10/19/9295905/premier-league-projections-and-new-expected-goals

Eggels, H., van Elk, R., & Pechenizkiy, M. (2016). Explaining soccer match outcomes with goal scoring opportunities predictive analytics. *CEUR Workshop Proceedings*.

Ensum, J., Pollard, R., & Taylor, S. (2004). Applications of logistic regression to shot at goal in association football: Calculation of shot probabilities, quantification of factors and player/team. *Journal of Sports Sciences*, *22*, 504.

Gauriot, R., & Page, L. (2017). Gauriot, romain, and lionel page. fooled by performance randomness: over-rewarding luck. *Working Paper*.

Gurpinar-Morgan, W. (2015, December). *Unexpected goals.* Retrieved from https://2plus2equals11.com/2015/12/31/unexpected-goals/

Henderson, A. D., Raynor, M. E., & Ahmed, M. (2012). How long must a firm be great to rule out chance? Benchmarking sustained superior performance without being fooled by randomness. *Strategic Management Journal*, *33*(4), 387–406.

Henry, N. (2017, February 12). *Rio Ferdinand: Manchester United deserve to be sixth in the Premier League.* Retrieved from http://www.squawka.com/news/rio-man-utd-deserve-to-be-sixth/900917#MpZwzGwgpjRJMyAC.97

IJtsma, S. (2015, August). *A close look at my new expected goals model.* Retrieved from http://11tegen11.net/2015/08/14/a-close-look-at-my-new-expected-goals-model/

Kwiatkowski, M. (2018, April). *Beyond the baseline: individual goalscoring, shotstopping and managerial skill.* Retrieved from https://www.optasportspro.com/de/ueber-uns/optapro-blog/posts/2018/video-ein-vortrag-von-marek-kwiatkowski-auf-dem-optapro-analytics-forum-2018

Lefgren, L., Platt, B., & Price, J. (2015). Sticking with what (barely) worked: A test of outcome bias. *Management Science*, *61*(5), 1121–1136.

Lepkowski, C. (2014, April). *Real Betis: We were wrong to let Pepe Mel go.* Retrieved from http://www.birminghammail.co.uk/sport/football/football-news/west-brom-real-betis-admit-7002341

Lucey, P., Bialkowski, A., Monfort, M., Carr, P., & Matthews, I. (2015). Quality vs. quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. In *Proc. 8th annual mit sloan sports analytics conference* (pp. 1–9).

NY Times. (2017, February). *How Arsenal and Arsene Wenger bought into analytics.* Retrieved from https://www.nytimes.com/2017/02/03/sports/soccer/arsenal-arsene-wenger-analytics.html

OptaPro. (2017a). *Advanced metrics: Expected goals.* Retrieved from http://optasports.com/news-area/news-expected-goals.aspx

OptaPro. (2017b). *BLOG: Expected goals in context.* Retrieved from http://www.optasportspro.com/about/optapro-blog/posts/2017/blog-expected-goals-explained/

Pollard, R., Ensum, J., & Taylor, S. (2004). Estimating the probability of a shot resulting in a goal: the effects of distance, angle and space. *International Journal of Soccer and Science*, *2*(1), 50–55.

Pollard, R., & Reep, C. (1997). Measuring the effectiveness of playing strategies at soccer. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *46*(4), 541–550.

Prozone Sports. (2015). *Performance.LAB innovation seminar: Tactical profiling.* Retrieved from https://vimeo.com/124518723

Raman, A. (2017, August). *How in the world does Nice overperform expected goals? Part two: The attack.* Retrieved from https://chanceanalytics.com/2017/09/24/how-in-the-world-does-nice-overperform-expected-goals-part-two-the-attack/

Rathke, A. (2017). An examination of expected goals and shot efficiency in soccer. *Journal of Human Sport and Exercise*, *12*, 514-529.

Stratabet. (2017, November). *StrataData - Differences that provide an edge.* Retrieved from https://www.stratabet.com/2017/11/27/stratadata-differences-provide-edge/

Taleb, N. (2005). *Fooled by randomness: The hidden role of chance in life and in the markets* (Vol. 1). Random House Incorporated.

Talksport. (2015, July). *Swansea City manager Garry Monk signs new contract.* Retrieved from https://talksport.com/football/swansea-city-manager-garry-monk-signs-new-contract-150710155912

The Guardian. (2014, December). *Cagliari sack manager Zdenek Zeman after one win in 12 matches.* Retrieved from https://www.theguardian.com/football/2014/dec/23/cagliari-sack-manager-zdenek-zeman

The Guardian. (2015a, December). *Garry Monk sacked by Swansea after one win in past 11 games.* Retrieved from https://www.theguardian.com/football/2015/dec/09/garry-monk-sacked-swansea

The Guardian. (2015b, July). *How Midtjylland took the analytical route towards the Champions League.* Retrieved from https://www.theguardian.com/football/

2015/jul/27/how-fc-midtjylland-analytical-route-champions-league
-brentford-matthew-benham

Thompson, S. (1999). Illusion of control: how we overestimate our personal influence. *Current Directions in Psychological Science*, *8*(6), 187–190.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*(4157), 1124–1131.

VfB Stuttgart. (2015, November). *VfB beendet Zusammenarbeit mit Alexander Zorniger.* Retrieved from http://www.vfb.de/de/vfb/aktuell/news-archiv/2015/308505-vfb-beendet-zusammenarbeit-mit-alexander-zorniger/?f3

Walerius, J. (2017, July). *Numbers aren't the problem (people are). The future of soccer analytics and what it might mean for the average fan.* Retrieved from https://fansided.com/2017/07/26/numbers-problem-soccer-analytics-future/

Wright, C., Atkins, S., Polman, R., Jones, B., & Sargeson, L. (2011). Factors associated with goals and goal scoring opportunities in professional soccer. *International Journal of Performance Analysis in Sport*, *11*(3), 438–449.